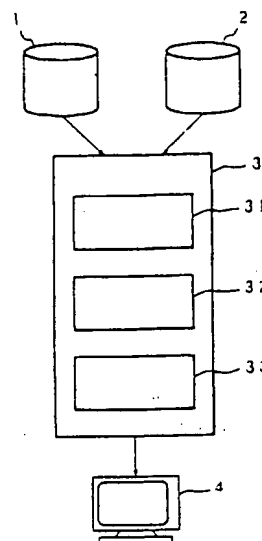


(54) DATA PROCESSOR

(11) 5-89171 (A) (43) 9.4.1993 (19) JP  
(21) Appl. No. 3-276887 (22) 28.9.1991  
(71) SUMITOMO METAL IND LTD (72) KAZUHIKO TSUDA(1)  
(51) Int. Cl<sup>5</sup>. G06F15/40, G06F15/20

**PURPOSE:** To speed up retrieval speed by reducing the number of the read times of characters from a document file and the number of comparison times with a character string being the object of retrieval.

**CONSTITUTION:** At the time of retrieving the character string from document data, a stage transition data generation means 31 generates state transition data where the transition of an inter-character state in the character string, which is stored in a retrieval object character string file 2, is defined. A character string detection means 33 moves a retrieval pointer which is set by a retrieval pointer setting means 32 and which moves on state transition data in accordance with the character which is read from the document data file 1, and detects the character string satisfying defined state transition as the retrieval object character string.



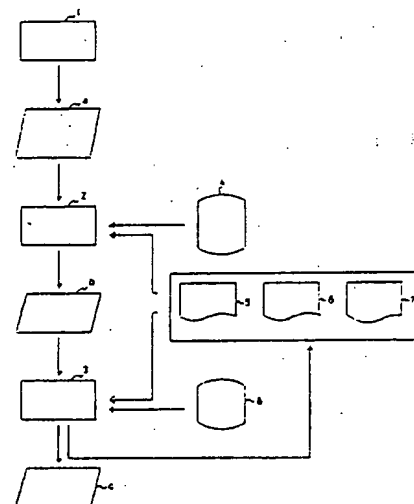
3: character string retrieval part

(54) RETRIEVAL DEVICE

(11) 5-89172 (A) (43) 9.4.1993 (19) JP  
(21) Appl. No. 3-248012 (22) 26.9.1991  
(71) NEC SOFTWARE LTD (72) SATOSHI SHIMOGAWARA(1)  
(51) Int. Cl<sup>5</sup>. G06F15/40

**PURPOSE:** To execute automatic restriction at the time of continuing a retrieval processing by automatically judging which retrieval result that is previously executed is set to be an object for restriction retrieval from a designated retrieval type, a retrieval content level and a retrieval history so as to generate a retrieval expression.

**CONSTITUTION:** It is assumed that the retrieval request of "What is less than one hundred thousand yen?" is inputted at first. A retrieval object decision part 2 extracts cost from a dictionary 4 since the type of the retrieval object is "ten thousand yen", and decides an item to be the retrieval object as cost. Thus, the retrieval expression (cost is less than one hundred thousand yen) using conventional Boolean logic is generated and the retrieval processing is executed. When the retrieval processing normally terminate it is assumed that "Tokyo city" is inputted in the retrieval request next. The retrieval object decision part 2 decides the retrieval object item to the name of a prefecture. The retrieval processing is executed with the retrieval result of first time as the object. When it does not exist, the second name of the prefecture in a retrieval history table 6 is set.



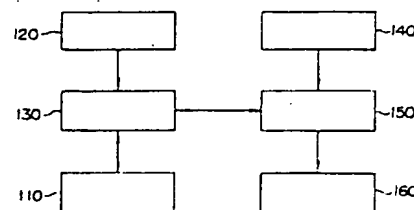
1: retrieval expression analysis part, 3: retrieval processing part, 5: retrieval content level table, 7: data base item pair corresponding table, 8: data base, a: comparison value of retrieval object and type of retrieval object, b: retrieval expression, c: retrieval result

(54) CLASSIFICATION DEVICE FOR STRUCTURED DOCUMENT

(11) 5-89173 (A) (43) 9.4.1993 (19) JP  
(21) Appl. No. 3-249580 (22) 27.9.1991  
(71) FUJI XEROX CO LTD (72) TOSHIAKI MORI  
(51) Int. Cl<sup>5</sup>. G06F15/40

**PURPOSE:** To easily classify a structured document in accordance with the purpose of a user.

**CONSTITUTION:** A retrieval condition designation means 120 designates structure element information as information on the retrieval of a document, and a classified attribute designation means 140 designates structure element information as information on the classification of the document. A structured document retrieval means 130 retrieves a structured document group based on structure element information as information on the retrieval of the document from plural structured documents stored in a structured document storage means 110 and transfers it to a structured document classification means 150. It classifies the transferred structured document group based on structure element information as information on the classification of the document. The classified result is displayed on a document display means 160.



(51)Int.Cl.<sup>5</sup>

G 0 6 F 15/40

識別記号

5 0 0 Q

庁内整理番号

7060-5L

T 7060-5L

F I

技術表示箇所

審査請求 未請求 請求項の数2(全 18 頁)

(21)出願番号 特願平3-249580

(22)出願日 平成3年(1991)9月27日

(71)出願人 000005496

富士ゼロックス株式会社

東京都港区赤坂三丁目3番5号

(72)発明者 森 俊明

神奈川県横浜市保土ヶ谷区神戸町134番地

横浜ビジネスパーク イーストタワー

富士ゼロックス株式会社内

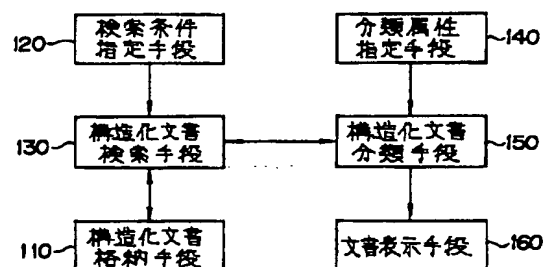
(74)代理人 弁理士 木村 高久

(54)【発明の名称】 構造化文書分類装置

(57)【要約】

【目的】構造化文書を、利用者の目的に応じて容易に分類することのできる構造化文書分類装置を提供すること。

【構成】検索条件指定手段120によって文書の検索に関する情報としての構造要素情報を指定し、また分類属性指定手段140によって文書の分類に関する情報としての構造要素情報を指定する。構造化文書検索手段130は、構造化文書格納手段110に格納されている複数の構造化文書の中から、文書の検索に関する情報としての構造要素情報に基づく構造化文書群を検索して、構造化文書分類手段150に転送する。構造化文書分類手段150では、その転送された構造化文書群を、文書の分類に関する情報としての構造要素情報に基づいて分類する。その分類結果は、文書表示手段160に表示される。



## 【特許請求の範囲】

【請求項1】文書の検索及び分類に関する情報として構造化文書を構成する構造要素に対応する構造要素情報を指定する指定手段と、

複数の構造化文書の中から、文書の検索に関する情報としての前記構造要素情報に基づく構造化文書群を検索する構造化文書検索手段と、

この構造化文書検索手段により検索された構造化文書群を、文書の分類に関する情報としての前記構造要素情報に基づいて分類する構造化文書分類手段とを具えたことを特徴とする構造化文書分類装置。

【請求項2】前記構造化文書分類手段は、前記構造化文書検索手段により検索された構造化文書群を分類した分類状態を表現している分類構造を生成することを特徴とする請求項1記載の構造化文書分類装置。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】この発明は構造化文書を検索し分類する構造化文書分類装置に関する。

## 【0002】

【従来の技術】大量の電子文書を格納／分類／整理するための装置として、現在ではファイルシステムや文書データベースが知られており、実際に多くのオフィスで利用されている。例えば、図18に示すように、大分類を「文書の種類」、小分類を「作成日」として、分類基準を決めて文書を整理する。

【0003】すると、「作成日」毎に分類して見たいときには「文書の種類」毎に全ての文書を見ていく必要がある。つまり、図18に示すように、「91年1月分」の文書をまとめて調べるためには「月報」「報告書」

「依頼書」の3箇所の「91年1月分」を調べなければならない。さらに、今までなかった分類基準である「作成者」毎に分類したいときには、初めから全ての文書を調べて分類し直す必要がある。

【0004】また文書データベース、電子ファイリングシステムでは、文書に対して、キーワードや属性を書き込める領域が用意されている。これらを利用することによって、文書を検索するという機能が付加された。この検索機能を利用すると、検索式を使って、必要な文書だけを取ってくることができるようになった。

【0005】上記装置以外の装置としては、特開昭63-157229号公報や特開平1-189721号公報に記載されたものが知られている。

【0006】特開昭63-157229号公報に記載されたものは、予め幾つかの分類をしておき、属性やキーワードを利用して分類を切り替えることによって、あたかも再分類したように再表示するものである。一方、特開平1-189721号公報に記載されたものでは、分類軸を予め複数用意して、1度に複数の分類軸を選択して多次元表示で文書の位置を表示するようにしている。

## 【0007】

【発明が解決しようとする課題】しかしながら、上記特開昭63-157229号公報や特開平1-189721号に記載されたものや、上記従来の文書データベース、電子ファイリングシステムには、次のような問題がある。

【0008】(1)文書に属性やキーワードを付与しなければならない。

【0009】多量の文書に属性やキーワードを付与するために大幅な手間がかかる。また属性やキーワードを付与しなければ再分類できない。

【0010】(2)文書の更新に対して、属性やキーワードの更新が伴わない。

【0011】文書の内容と属性とを別々に管理しているためである。これを回避する技術として文書の全文検索やキーワードの抽出などが一般に知られている。しかし、それでは検索対象として、目的に合わない無駄な文書(余計な部分の文書)を検索したり、検索時間がかかったりする。無駄な文書が多いほど、目的の文書を探すために手間がかかる。

【0012】(3)検索式を何度も書き直して検索しなければならない。

【0013】一般に、条件検索は属性値やキーワードがハッキリしているときには都合が良い。しかし、目的とする文書が属性やキーワードでは特定できないときや、初めて検索するときには、検索式の属性値やその範囲を変更し試行錯誤することになる。従って、何度も検索式を書き直さなければならず、大変な作業量／時間になる。

【0014】(4)分類項目どおりの分類を保守する手間がかかる。

【0015】文書を分類／整理するための分類項目の保守はユーザ任せである。例えば図19に示す例では、

「91年1月分の報告書」を探すときに「91年1月分」「報告書」のどちらを探したら良いかは、文書を格納した人にしか分からない。またある文書をその文書とは全く関係のないところに置くこともできてしまう。それらを保守するための作業量は大きなものである。

【0016】(5)文書を予め分類する手間がかかる。

【0017】特開昭63-157229号公報や特開平1-189721号公報に記載されたものでは、分類項目を「分類の種類」や「作成日」と変えて見ることができるが、そのためには予め必要とする分類項目で全ての文書を分類しておく必要がある。

【0018】そこで、本発明は、構造化文書を、利用者の目的に応じて容易に分類することのできる構造化文書分類装置を提供することを目的とする。

## 【0019】

【課題を解決するための手段】上記目的を達成するため、この発明では、文書の検索及び分類に関する情報と

して構造化文書を構成する構造要素に対応する構造要素情報を指定する指定手段と、複数の構造化文書の中から、文書の検索に関する情報としての前記構造要素情報に基づく構造化文書群を検索する構造化文書検索手段と、この構造化文書検索手段により検索された構造化文書群を、文書の分類に関する情報としての前記構造要素情報に基づいて分類する構造化文書分類手段とを具えている。

#### 【0020】

【作用】この発明によれば、指定手段によって、文書の検索に関する情報としての構造要素情報と、文書の分類に関する情報としての構造要素情報とを指定すると、構造化文書検索手段は、複数の構造化文書の中から、文書の検索に関する情報としての構造要素情報に基づく構造化文書群を検索する。更に、その検索された構造化文書群を、構造化文書分類手段が、文書の分類に関する情報としての構造要素情報に基づいて分類する。

【0021】従って、構造化文書を構成する構造要素を、構造要素名を用いて指定することにより、リアルタイムで、複数の構造化文書を検索し、かつ、分類することができ。

#### 【0022】

【実施例】以下、本発明の実施例を添付図面を参照して説明する。

【0023】最初に、本発明に係る構造化文書分類装置の構成の説明に先立って、本発明に係る構造化文書の検索のモデルを説明する。

【0024】図2は、本発明に係る構造化文書の検索のモデルを図示化したものであり、同図において、検索フォルダ210は検索式を持つフォルダであり、その検索式は「検索対象」、「検索条件」、「分類属性」の組みから構成されている。「検索対象」はフォルダ名であり、フォルダ内の全ての文書を検索対象としていることを意味する。「検索条件」は論理式であり、「構造要素名 演算子 値」を用いて記述する。「分類属性」は構造要素名であり、対応する構造化文書の構造要素の内容別に分類する。

【0025】例えば、検索対象220で表されるフォルダ名が「検索対象」という検索対象フォルダ230に存在する文書集合中、この場合は、検索対象フォルダ230内に存在する文書集合231、検索フォルダ230に存在するフォルダ232の下位に存在する文書集合233、234の各文書集合中、検索条件240に適合している文書だけが検索（選択）され、更に検索された文書は、分類属性250で表される構造要素の内容に応じて部分集合260、270に分類される。この分類された部分集合はそれぞれ仮想フォルダ280、290内に表示される。仮想フォルダ280、290はフォルダの実体が存在せず、検索／分類結果をフォルダの形で表現したビューである。

【0026】すなわち、この実施例では、検索フォルダ210の検索式を評価することで、検索フォルダ210の下位に、検索及び分類結果としての部分集合（文書集合）260、270を見せるためのビューである仮想フォルダ280、290が生成される。このため、検索及び分類結果をユーザに提示するためのデータ構造は、図2に示すように、根（ルート）である検索フォルダの下位に仮想フォルダが位置し、その仮想フォルダの下位に文書が位置している構造（以下、この実施例では分類構造という）になっている。なお、検索フォルダ、仮想フォルダはそれぞれ文書を分類して表示するためのオブジェクトである。このため、表示方式が変更されれば他のオブジェクトを導入することになる。

【0027】ここで、検索フォルダと仮想フォルダとの違いは、前者はユーザが定義／作成する実態であり、後者は検索毎に更新されるビューである。また前者は分類構造の根であり、後者は分類構造の中間位置に存在する（分類構造の末端は文書である）。

【0028】なお、従来のファイルシステムとのインタフェースを同じにするために、ここでは分類の道具として「仮想フォルダ」を導入している。この「仮想フォルダ」とは、検索式を持ち、その検索条件に合う文書集合があたかもその仮想フォルダの中にあるように見せるものである。仮想フォルダを生成することによって、ユーザの目的に合わせた一定の条件に合う文書の分類を実現することができる。このように本発明では、従来の階層ファイルシステム上にユーザの目的に合わせて多量の文書を自動的に分類してユーザに見せることができるようにしたものである。

【0029】図1は発明に係る構造化文書分類装置の第1の実施例を機能ブロック図で示したものである。

【0030】同図において、構造化文書格納手段110には文書（構造化文書）や複数の文書を所有するフォルダが格納される。構造化文書格納手段110からは、格納された文書やフォルダを識別子や名前を取り出すことができる。

【0031】検索条件指定手段120は、検索時に必要な情報、具体的には、構造化文書格納手段110に格納された複数の構造化文書やフォルダ（検索対象）と検索条件とを指定する（上記検索式内の「検索対象」と「検索条件」とを指定する）。その指定され情報は、構造化文書検索手段130に入力される。

【0032】構造化文書検索手段130は、構造化文書格納手段110内から、検索条件指定手段120で指定された文書の内容を検索して、検索条件に合う文書を検索する。例えば図2においては、検索対象220で指定された検索対象フォルダ230内に存在する文書集合231、233、234から、検索条件240に適合する部分集合260、270を有する部分集合群（この場合は、図2に示す部分集合260、270のように分類さ

れておらず、ランダムな文書の集合になっている)を検索する。

【0033】分類属性指定手段140は、分類に必要な情報である構造要素名(分類属性)を指定する(上記検索式内の「分類属性」を指定する)。その指定された情報は構造化文書分類手段150に入力される。なお、分類属性の指定は1つであっても良いし、複数であっても良い。

【0034】ここで1つの分類属性が指定された場合は、1つの構造要素に注目して、それを分類属性として、検索時に文書集合を分類して、分類構造を生成する。

【0035】また複数の分類属性が指定された場合は、複数の構造要素名に対応する構造要素に対応付けられた文書の内容による分類を繰り返し、階層的な部分構造を生成する。つまり、複数存在する分類属性を順に1つずつ評価して再帰的に分類を続けて、分類構造を階層化して生成する。

【0036】なお、分類構造を生成しなければ、分類された文書を表示できないというものではないので、分類構造を生成しなくても良い。しかし、この実施例では、分類構造を生成するものとして説明する。なお分類構造を生成しない場合の表示については後述する。

【0037】構造化文書分類手段150は、構造化文書検索手段130によって検索された文書を、分類属性指定手段140で指定された構造要素名に対応付けられた文書の内容別に分類するとともに、その分類された文書を基に分類構造を生成する。例えば図2においては、分類属性250に従って、部分集合群を、部分集合260と、部分集合270とに分類するとともに、仮想フォルダ280、290を生成して分類構造を生成する。

【0038】文書表示手段160は、構造化文書分類手段150によって生成された分類構造、構造化文書検索手段130によって検索された文書の一部又は全部を表示する。

【0039】次に上述した各構成要素を具体的に説明する。構造化文書格納手段110は、構造化文書を格納するためのファイルシステムやファイルサーバ、データベースの機能を有する磁気ディスク装置や光磁気ディスク装置など記憶装置で実現される。検索条件指定手段120および分類属性指定手段140は、マウスなどのポインティングデバイスやキーボードで実現される。文書表示手段160はディスプレイ装置や印字装置など出力装置で実現される。

【0040】構造化文書検索手段130は、上述した当該検索手段130の機能を遂行させるためのプログラム、即ち、格納された構造化文書を構成する各構造要素を、指定された構造要素名で識別し、その構造要素名に対応する文書内容を検索するためのプログラムを、プロセッサや中央処理装置等の制御手段(図示せず)が実行

することにより実現される。構造化文書分類手段150は、上述した当該分類手段150の機能を遂行させるためのプログラム、即ち、構造化文書検索手段130による検索結果としての、文書構造と文書内容とを参照して、その文書内容によって文書を分類するためのプログラムを、上記制御手段が実行することにより実現される。

【0041】なお、図示しない上記制御手段は、構造化文書格納手段110に対するアクセス制御、検索条件指定手段120および分類属性指定手段140に対する入力制御、文書表示手段160に対する出力制御を行っている。

【0042】この実施例では、検索/分類に関する操作はデスクトップ環境で行えるようになっており、アイコン表示されたフォルダ(フォルダアイコン)をマウスボタンでダブル・クリックすると、そのフォルダアイコンに対応するフォルダ内の文書や他のフォルダ(例えば仮想フォルダ)を見ることができる。

【0043】検索フォルダには、予めプロパティ・キーを押下してプロパティ・シートを開き、そのプロパティ・シートで検索式を設定しておく。その検索式を設定するのが、検索条件指定手段120、分類属性指定手段140である。このとき、ダブル・クリックすると、条件選択された文書や生成された仮想フォルダを見ることができる。なお、検索フォルダに予め検索式が設定されていないときは、検索直前にプロパティ・シートが開くだけであり、検索は行われない。従ってシステムはユーザに対して検索式の入力を要求することとなる。

【0044】上述した構成において、構造化文書分類装置の処理動作を図3のフローチャートを参照して説明する。

【0045】最初に、検索条件指定手段110によって、文書集合(検索対象)および検索条件を入力するとともに、分類属性指定手段140によって、構造要素名(分類属性)を入力することにより、検索/分類に必要な検索式を入力する(ステップ301)。その検索対象及び検索条件は構造化文書検索手段130に入力され、構造要素名(分類属性)は構造化文書分類手段150に入力される。

【0046】すると、構造化文書検索手段130は、得られた検索対象及び検索条件を評価し実行する。すなわち構造化文書格納手段110に格納されている検索対象の文書集合から、検索条件に適合する文書を検索して(ステップ302)、その検索結果としての文書を構造化文書分類手段150に渡す。

【0047】一方、構造化文書分類手段150は、検索式に分類属性が指定されているか否かを判断する(ステップ303)。ここで、分類属性(構造要素名)が指定されていれば、その分類属性と、検索された文書集合中の当該分類属性に対応する文書内容とに基づいて、その

文書集合を分類し（ステップ304）、その後、分類構造を生成する（ステップ305）。更に、分類された文書集合を分類構造として文書表示手段160に出力する（ステップ306）。なお、ステップ303において分類属性が指定されていない場合は、検索された文書を文書表示手段160に出力する。

【0048】次に上述した構造化文書の検索／分類処理を、図4乃至図15を用いて具体的に説明する。

【0049】図4は構造化文書格納手段110に格納されている構造化文書の一例を示したものであり、この構造化文書は、文書の論理的な構造を形成する各構造要素と、該構造要素に対応付けられた文書の内容及び構造要素名と、該構造要素から構成される論理オブジェクトとから構成されている。図4において401～406は文書の論理的な構造を形成する各構造要素であり、特に401は「文書論理根」である。また402A～405Aは文書の内容（内容部）である。さらに「出願人」、「公開年」、「国際特許分類」、「タイトル」、「本文」は、それぞれ構造要素に対応付けられた構造要素名である。なお、「本文」の構造要素406以下の構造は複雑になっているので、ここではその構造を省略している。

【0050】ここで、「出願人」という構造要素名に対応付けられる文書内容としては、例えば「AA会社」、「BB会社」など出願人の名称であり、また「公開年」という構造要素名に対応付けられる文書内容としては、例えば「1987年」、「1988年」など出願公開年であり、更に「国際特許分類」という構造要素名に対応付けられる文書内容としては、例えば「G06F 15/20」、「G06F 7/00」など国際特許分類（IPC）である。

【0051】この実施例では、少なくとも、「出願人」、「公開年」、「国際特許分類」の各構造要素を含んでいる構造化文書が、例えば「日米特許フォルダ」（実際にはファイルである）内に複数格納されているものとする。勿論、他の構造化文書も格納されていても良い。

【0052】図5は、「検索対象」、「検索条件」、「分類属性」からなる検索式を指定するための検索フォルダのプロパティ・シート500を示したものであり、プロパティ・シート500には、構造化文書を検索するために必要な「検索対象」および「検索条件（式）」を指定するための各フィールド510、520と、文書を自動的に分類するための分類項目として「分類属性」を指定するためのフィールド530とが設けられており、更に「検索条件」に対応して、複数の条件式が指定された場合に、その複数の条件式の「論理和」かあるいは「論理積」のいずれかを選択するボタン520Aが設けられている（これは、条件式は複雑な式を許しても良いが、ユーザ・インタフェースを簡単にするために簡単な表現形式にしている。）ここで、「検索対象」の指定に際して

は、フォルダ名をフィールド510に指定する。これは、指定されたフォルダ名に対応するフォルダ内にある全ての文書を要素とする文書集合が検索対象となることを意味している。また「検索条件」の指定に際しては、「検索対象」で示された文書集合のうち、目的の文書に必要な条件式を、フィールド520に記述する。複数の条件式を設定した場合には、ボタン520A内の「AND」あるいは「OR」のいずれかの領域上にカーソルを移動してマウスボタンをクリックすることにより、「論理和」かあるいは「論理積」のいずれかを選択する。

【0053】上記条件式のフォーマットは「（構造要素名 演算子 値）」である。条件式の評価は、「条件式の構造要素名」に文書の構造要素の内容が代入されて演算子が評価される。条件式は論理式であり、真となった文書が選択される。

【0054】最後に「分類属性」の指定に際しては、「検索対象」と「検索条件」とを指定することにより得られた文書集合を分類するために必要な分類項目、つまり分類属性を、フィールド530に記述する。「分類属性」で指定される構造要素の内容別に文書が分類され、部分集合に分けられる。更にその部分集合に対応する仮想フォルダが自動的に作成される。

【0055】ここで、ユーザが、図5に示すように、フィールド510内にフォルダ名として「日米特許フォルダ」を指定し、フィールド520内に条件式として（国際特許分類=G06F 15/20）を記述し、またボタン520Aの「OR」を指示し、更にフィールド530内に構造要素名（分類属性）として「出願人」を記述したとする。

【0056】このようにして検索式が指定されると、構造化文書検索手段130は、フィールド510内に指定されたフォルダ名に対応する「日米特許フォルダ」を構造化文書格納手段110から検索するとともに、その「日米特許フォルダ」内の構造化文書集合から、フィールド520内に指定された条件式「国際特許分類=G06F 15/20」に適合する文書を検索する。このとき、指定された条件式「国際特許分類=G06F 15/20」と同一の構造要素名に対応する構造要素を有している構造化文書のみが検索される。例えば、図4に示す構造化文書の構造要素404に対応する文書内容404Aを参照することで、当該文書が、検索対象の文書であるか否かを判定することができる。ここでは複数の文書が検索されるものとする。その検索された各構造化文書は構造化文書分類手段150に伝送される。

【0057】すると、構造化文書分類手段150は、フィールド530に指定された「出願人」という分類属性に従って渡された各構造化文書を分類して、各部分毎に部分集合にする。このとき、指定された「出願人」と同一の構造要素名で表された構造要素に対応する文書内容に従って分類することになる。例えば、図4に示す構造化

化文書の構造要素402に対応する文書内容402Aを参照することにより、細分類することができる。又、分類された各部分集合に対応して仮想フォルダを生成し、その後、部分集合と仮想フォルダとに基づいて分類構造を生成する。その分類構造の一例を図6に示す。

【0058】図6に示す例では、ルートである検索フォルダ610の下位に「AA会社」、「BB会社」、「CC会社」の各出願人毎にそれぞれ対応して分類された各部分集合620A、630A、640Aを表示すべくビューとしての各仮想フォルダ620、630、640が生成されている。

【0059】この図に示した状態が内部状態であって、検索／分類処理が終了した状態でもある。このとき、文書表示手段160には、例えば、検索フォルダ610に対応するウィンドウ内に、例えば、仮想フォルダ620、630、640をそれぞれ示す各仮想フォルダアイコンと、対応する出願人「AA会社」、「BB会社」、「CC会社」とが表示されることとなる。

【0060】そして、ユーザが、例えば出願人「CC会社」に対応する仮想フォルダアイコンをマウスボタンをクリックして、仮想フォルダのプロパティ・シートを表示すべく指示をすることにより、出願人「CC会社」に対応して分類された部分集合が得られるための、検索対象と検索条件とが記述された仮想フォルダのプロパティ・シートが文書表示手段160に表示される。その表示状態の一例を図7に示す。

【0061】この図では、検索／分類した文書集合から作成された仮想フォルダ（ここでは検索フォルダ（／動的分類フォルダ）の下位に、仮想フォルダ（／動的分類フォルダ／CC会社）が生成された仮想フォルダ）のプロパティ・シート（検索式）700を示している。検索対象が検索フォルダ（／動的分類フォルダ）になり、検索条件が「出願人＝CC会社」となる。

【0062】上述した例では分類属性指定手段140による分類属性の指定が1つであったが、複数の分類属性を指定することもできる。

【0063】これには、例えば検索フォルダと仮想フォルダの区別なく、検索式中に「分類属性」を書けるようにすると、複数の分類属性が書けるようになり、分類したものを更に細分類することができる。このようにすると、分類の階層を動的に生成して、分類構造を実現できる。次に複数の分類属性を指定した場合の検索／分類について説明する。

【0064】図8は、複数の分類属性を指定した場合の検索フォルダのプロパティ・シートの一例を示したものである。プロパティ・シート800は、図5に示したプロパティ・シート500と同一の構成になっている。図8に示した例では、分類属性に対応するフィールド530内には、分類属性として「出願人」、「公開年」がそれぞれ指定されている。

【0065】この検索式が指定されると、構造化文書検索手段130は、上述したように構造化文書格納手段110内に格納されている「日米特許フォルダ」内の構造化文書集合から、フィールド520内に指定された条件式「国際特許分類＝G06F 15/20」に適合する文書集合を検索する。その文書集合は構造化文書分類手段150に通知される。

【0066】すると、構造化文書分類手段150は、フィールド530に指定された「出願人」及び「公開年」という分類属性に従って、渡された各構造化文書を分類して各部分集合にする。

【0067】このとき、検索された構造化文書群が、指定された「出願人」と同一の構造要素名で表された構造要素に対応する文書内容に従って分類され、更に、その出願人毎に分類された文書集合が、指定された「公開年」と同一の構造要素名で表された構造要素に対応する文書内容に従って分類されることとなる。最終的に分類された文書集合が部分集合となる。

【0068】次に、構造化文書分類手段150は、分類された各部分集合に対応して、出願人毎に分類した内容を表示するための仮想フォルダと、公開年毎に分類した内容を表示するための仮想フォルダとを生成し、その後、部分集合と各仮想フォルダとに基づいて分類構造を生成する。その分類構造の一例を図9に示す。

【0069】図9に示す例では、ルートである検索フォルダ910の下位に仮想フォルダ920、930、940が生成されている。仮想フォルダ920の下位に、部分集合921A及び部分集合922Aをそれぞれ表示すべくビューとしての、仮想フォルダ621及び仮想フォルダ622が生成されている。また仮想フォルダ930の下位に、部分集合931A及び部分集合932Aをそれぞれ表示すべくビューとしての、仮想フォルダ631及び仮想フォルダ632が生成されている。更に、仮想フォルダ940の下位に、部分集合941A～944Aをそれぞれ表示すべくビューとしての、仮想フォルダ941～944が生成されている。

【0070】このようにして検索／分類が終了すると、文書表示手段160には、上述したように仮想フォルダ920、930、940をそれぞれ示す各仮想フォルダアイコンが表示され、更に、それらの仮想フォルダアイコンの中から所望のアイコン例えば、仮想フォルダ940に対応するアイコンを指示すると、今度は仮想フォルダ941～944の各仮想フォルダにそれぞれ対応する仮想フォルダアイコンが表示されることとなる。

【0071】ここで、もう少し具体的に説明すると、検索フォルダ（／動的分類フォルダ）をダブル・クリックすると、ウィンドウが開いて、出願人毎に分類された部分集合を表示すべく仮想フォルダに対応する仮想フォルダアイコンが表示される。その表示状態の一例を図10に示す。



【0072】更に、例えば出願人「CC会社」に対応する仮想フォルダ（／動的分類フォルダ／CC会社）に対応する仮想フォルダアイコン1000をダブル・クリックすると、公開年毎に分類された部分集合を表示すべく仮想フォルダに対応する仮想フォルダアイコンが表示される。その表示状態の一例を図11に示す。

【0073】そして、例えば「1987年」に対応する仮想フォルダアイコン1100をダブルクリックすると、出願人が「CC会社」であって、公開年が「1987年」である部分集合（つまり構造化文書）のリストが文書表示手段160の表示画面上に表示される。そのリストの中から所望の構造化文書の項目を指定することにより、その文書の内容が表示されることとなる。

【0074】また構造化文書がどのようにして分類されたかを知るためには、仮想フォルダアイコンをダブル・クリックして、仮想フォルダのプロパティ・シートをオープンすべき旨の指示をすれば良い。例えば図10に示す「CC会社」の仮想フォルダアイコン1000をダブル・クリックして、仮想フォルダのプロパティ・シートをオープンすべき旨の指示をすることにより、細分類可能な仮想フォルダのプロパティ・シートが表示される。その表示状態を図12に示す。プロパティ・シート1200からは、検索対象として「検索フォルダ（動的分類フォルダ）」が、検索条件として「出願人＝CC会社」が、分類属性として「公開年」がそれぞれ指定されていることが分かる。

【0075】また、図11に示す「1987年」の仮想フォルダアイコン1100をダブルクリックして、仮想フォルダのプロパティ・シートをオープンすべき指示をすることにより、もう分類しない仮想フォルダのプロパティ・シートが表示される。その表示状態の一例を図13に示す。プロパティ・シート1300からは、検索対象として「仮想フォルダ（／動的分類フォルダ／CC会社）」が、検索条件として「公開年＝1987年」がそれぞれ指定されていることが分かる。

【0076】上述したように階層的な関係になっている仮想フォルダを順次指示してオープンすることにより、目的とする文書の内容を見ることが出来る。

【0077】しかしこれでは仮想フォルダの末端までいかなければ、目的の文書を見ることができない。そこで、表示の形式を変えて、検索フォルダに対応する検索フォルダアイコンをダブル・クリックすることにより、下部に文書集合をリストして、上部にフォルダの階層構造（分類構造）を表示できるようなウィンドウを表示する。その表示状態の一例を図14に示す。この図では、ウィンドウ1400のリスト表示領域1401内に、動的分類フォルダアイコン1402に対応する動的分類フォルダ（検索フォルダ）内に存在する文書のリストが表示されている。

【0078】この表示状態から、分類／検索結果を表示

すべく指示をすると、図15のようになる。この図では、ウィンドウ1400の階層構造表示領域1501内に、階層構造（分類構造）が表示されている。

【0079】例えば、フォルダアイコン1402は図9に示す検索フォルダ910に、フォルダアイコン1502は図9に示す仮想フォルダ930に、フォルダアイコン1503は図9に示す仮想フォルダ940に、フォルダアイコン1504、1505は図9に示す仮想フォルダ931、932に、フォルダアイコン1506、1507は図9に示す仮想フォルダ941、942に、それぞれ対応するものである。

【0080】次に、分類構造を生成せずに、検索フォルダ（動的分類フォルダ）用のウィンドウ内に、分類された文書を表示し、かつ、分類に併せて仕切りを表示するようにした表示例を図16に示す。この図の例に示すように、ウィンドウ1600内には、「AA会社」、「BB会社」、「CC会社」の各出願人にそれぞれ対応して、出願人毎に分類された文書のリストが表示されている。

【0081】図17は、発明に係る構造化文書分類装置の第2の実施例を機能ブロック図で示したものである。

【0082】この機能ブロック図は、図1に示した第1の実施例の機能ブロック図の構成に文書分類構造管理手段1700を追加した構成になっている。なお、同図において、図1に示した構成要素と同一の機能を果たす部分には同一の符号を付している。

【0083】この第2の実施例の装置は、基本的には第1の実施例の装置と同様の機能を有している。

【0084】文書分類構造管理手段1700には、最低文書数が記述されている。文書分類構造管理手段1700は、実際には、構造化文書分類手段150が必要とする細かな設定や暗黙の設定を記述したデータファイルである。そこには例えばソートの条件や検索／分類の実行タイミングなどが記述される。

【0085】これによって、例えば、文書が既に少なくなっているために分類する必要がない場合は、文書が少なくなり選択した文書の数で最低文書数以下のときは、分類を停止して、これ以上、階層の末端には仮想フォルダを生成しない。という指定ができる。またフォルダ内の文書の見易さを考慮して、ある項目（構造要素名）を指定して、それに対応する文書内容に応じて文書集合をソートすることも挙げられる。これによって目的の文書を素早く得ることができる。

【0086】次に、文書分類構造管理手段1700内に記述される記述内容の具体例を幾つか挙げて説明する。

【0087】（1）「そのほか」の指定（その1）ある数より小さいときに、「そのほか」のフォルダを生成して、そこにまとめて分類する。その数を指定する。

【0088】文書分類構造管理手段1700に、例えば「そのほか：出願人：5」と指定すると、構造化文書分



類手段150は、出願人で分類するとき、例えばXX会社は4件しかなかったら「そのほか」に分類し、また例えばAA会社が7件あったならば仮想フォルダ「AA会社」に分類する。

【0089】(2)「そのほか」の指定(その2)  
ある属性値だけに注目して分類し、該当しない値はすべて「そのほか」に入れる。

【0090】文書分類構造管理手段1700に、例えば「そのほか：出願人：(AA会社、BB会社、CC会社)」と指定すると、構造化文書分類手段150は、出願人については、「AA会社」、「BB会社」、「CC会社」、「そのほか」の4つのどれかに分類する。

【0091】(3)複数を1つにまとめる指定  
複数の値を指定して同じフォルダに分類する。

【0092】文書分類構造管理手段1700に、例えば「出願人：FXグループ：FXXXXX, FXIIII」と指定すると、構造化文書分類手段150は、FXXXXX、FXIIIIのどちらの特許も「FXグループ」にあるように分類する。

【0093】(4)範囲を指定  
条件式で記述できる範囲を指定して同じフォルダに分類する。

【0094】文書分類構造管理手段1700に、例えば「公開年：89年以前：公開年<=89年」と指定すると、構造化文書分類手段150は、公開年で分類する際に、条件式「公開年<=89年」に合うものは優先的にフォルダ「89年以前」に分類する。また文書分類構造管理手段1700に、例えば「公開年：90年から91年：公開年=90年/公開年=91年」と指定すると、構造化文書分類手段150は、公開年で分類する際に、条件式「公開年=90年/公開年=91年」に合うものは優先的に「90年から91年」に分類する。このように条件式を複数指定することもできる。

【0095】次に第1及び第2の実施例における変形例(応用例)を説明する。

【0096】(1)検索式の再利用

一度指定された検索式を構造化文書格納手段110に格納しておき、その検索式を、指定手段(検索条件指定手段120、分類属性指定手段140)により呼び出すことにより、同一の検索式を再度入力しなくても良いようにする。これによりユーザが検索式を入力する手間が軽減される。

【0097】(2)指定手段の改善

「検索対象」の選択は、フォルダ名を指定するのではなく、フォルダアイコンをコピーする。また「検索条件」や「分類属性」の選択はメニューにすると、よりユーザ・インタフェースが良くなる。

【0098】(3)実行速度の短縮

文書が変更されていないときには、一度検索した分類については情報を格納して検索せずに再利用することによ

って、実行速度を短縮する。具体的には、検索式と分類構造と分類した時間とを格納して、文書がフォルダの変更時間と分類した時間とを比べて、文書に変更がなければ格納した分類構造を表示する。変更されているときには検索して分類し直す。

【0099】以上説明したように、本実施例によれば次のような利点がある。

【0100】(1)文書に属性やキーワードを付与しなくても分類できる。

【0101】構造化文書の構造要素を文書の属性のように扱うことで、キーワードをつけたり、属性を入力する必要がなく、文書を検索/分類することができる。

【0102】(2)検索に必要な属性と文書の一貫性が維持できる。

【0103】構造化文書の構造要素を文書の属性のように扱うことで、文書内容の更新に合わせて、検索/分類属性との一貫性が維持できる。

【0104】(3)検索式の更新が最小限で済む。

【0105】本発明は単なる検索で目的の文書集合を選択するだけでなく、選択された文書集合を細かく分類するようにしたものである。そこで、検索条件で大まかな文書集合を選択して、これを分類項目で分類する検索式を設定しておく。すると、検索式をいちいち書いたり、書き直したりしないで、多目的に検索式を使うことができる。

【0106】(4)文書集合を分類項目に沿って多段に細分類できる。

【0107】分類項目を多数指定することによって、単なる分類だけでなく、多段に分類することで、細分類ができる。多段に分類することによって、ユーザ自身の分類構造を持つことができる。とくに複数のユーザが共有している文書を探し出すときには、他のユーザが利用している分類構造/分類基準にとらわれずに、同じ文書集合を自分の分類構造で分類してみることができる。

【0108】(5)文書の分類はその場で自動的に行う。

【0109】特開昭63-157229号公報や特開平1-189721号公報に記載されたもののように予め文書を分類する必要がなく、ユーザは「分類属性」の指定だけで、本実施例の装置が全て自動的に分類する。

【0110】最後に、本発明の特徴(新規の技術)を箇条書きにしてまとめる。

【0111】(1)その場で複数の分類構造を生成できる。

【0112】(2)複数の分類をその場で重ねて階層的な1つの構造として見せること。

【0113】(3)文書の構造要素そのものが属性となること。

【0114】(4)検索式を利用して、漠然とした文書集合を取り出して分類すること。

【0115】

【発明の効果】以上説明したように本発明によれば、構造化文書の検索及び分類に必要な情報として、構造化文書を構成する構成要素を指定することにより、検索条件として指定された構成要素を有している構造化文書群が検索され、更に、その検索された構造化文書群が、分類属性として指定された構成要素を有している構造化文書毎に分類されることとなり、従来の如く、複数の構造で文書を予め分類しておく必要はなく、文書の検索／分類操作指示により、リアルタイムで複数の分類構造を生成でき、かつ、表示することができる。

【図面の簡単な説明】

【図1】本発明に係る構造化文書分類装置の第1の実施例を示す機能ブロック図。

【図2】本発明に係る構造化文書の検索及び分類のモデルを図示化した図。

【図3】本実施例における検索及び分類の処理動作を示すフローチャート。

【図4】本実施例における構造化文書格納手段に格納される構造化文書の一例を示す図。

【図5】本実施例における検索式を指定するための検索フォルダのプロパティ・シートの一例を示す図。

【図6】図5に示した検索式に基づいて生成された分類構造の一例を示す図。

【図7】本実施例におけるビューとしての仮想フォルダのプロパティ・シートの一例を示す図。

【図8】本実施例における複数の分類属性を有する検索

式を指定するための検索フォルダのプロパティ・シートの一例を示す図。

【図9】図8に示した検索式に基づいて生成された分類構造の一例を示す図。

【図10】本実施例における検索フォルダのビューの一例を示す図。

【図11】本実施例における仮想フォルダのビューの一例を示す図。

【図12】本実施例における細分類可能な仮想フォルダのプロパティ・シートの一例を示す図。

【図13】本実施例における分類しない仮想フォルダのプロパティ・シートの一例を示す図。

【図14】本実施例における検索／分類前の検索フォルダの表示状態の一例を示す図。

【図15】本実施例における検索／分類後の検索フォルダの表示状態の一例を示す図。

【図16】本実施例における仮想フォルダの無い検索フォルダを示す図。

【図17】本発明に係る構造化文書分類装置の第2の実施例を示す機能ブロック図。

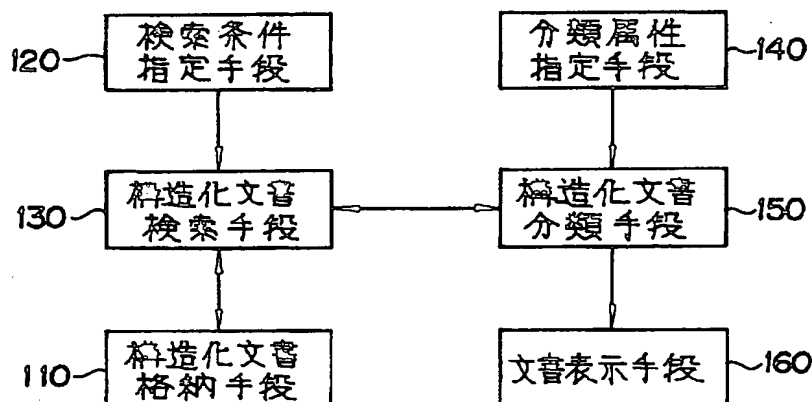
【図18】複数の文書の分類の一例を示す図。

【図19】複数の文書の分類の一例を示す図。

【符号の説明】

110…構造化文書格納手段、120…検索条件指定手段、130…構造化文書検索手段、140…分類属性指定手段、150…構造化文書分類手段、160…文書表示手段、1700…文書分類構造管理手段。

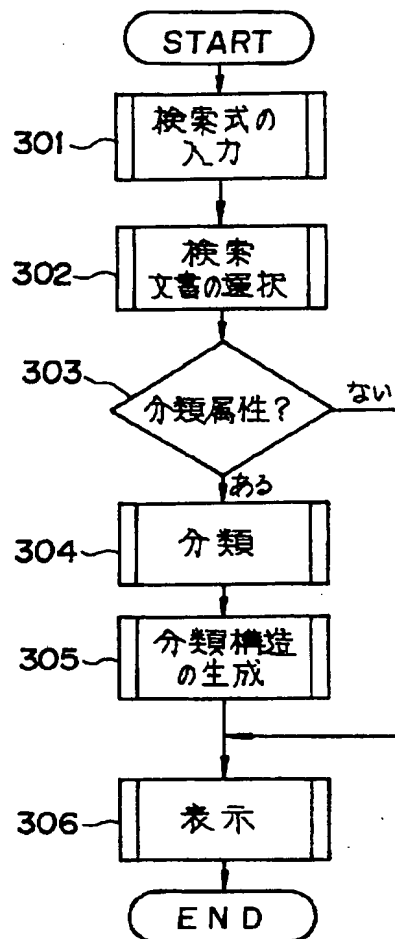
【図1】



```

graph TD
    401[文書論理根] --- 402[出願人]
    401 --- 403[公開年]
    401 --- 404[IPC]
    401 --- 405[タイトル]
    401 --- 406[本文]
    402 --- 402A[ ]
    403 --- 403A[ ]
    404 --- 404A[ ]
    405 --- 405A[ ]
  
```

【図3】



【図7】

700 プロパティシート

Profile : /動的分類フォルダ/CC会社

検索対象 : /動的分類フォルダ

検索条件 : (出願人 = CC会社 )

【図12】

1200 プロパティシート

Profile : /動的分類フォルダ/CC会社

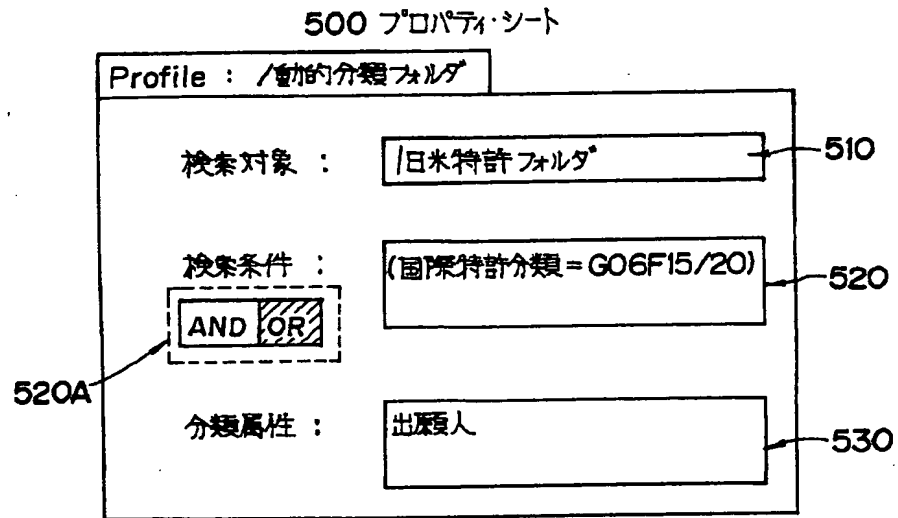
検索対象 : /動的分類フォルダ

検索条件 : (出願人 = CC会社 )

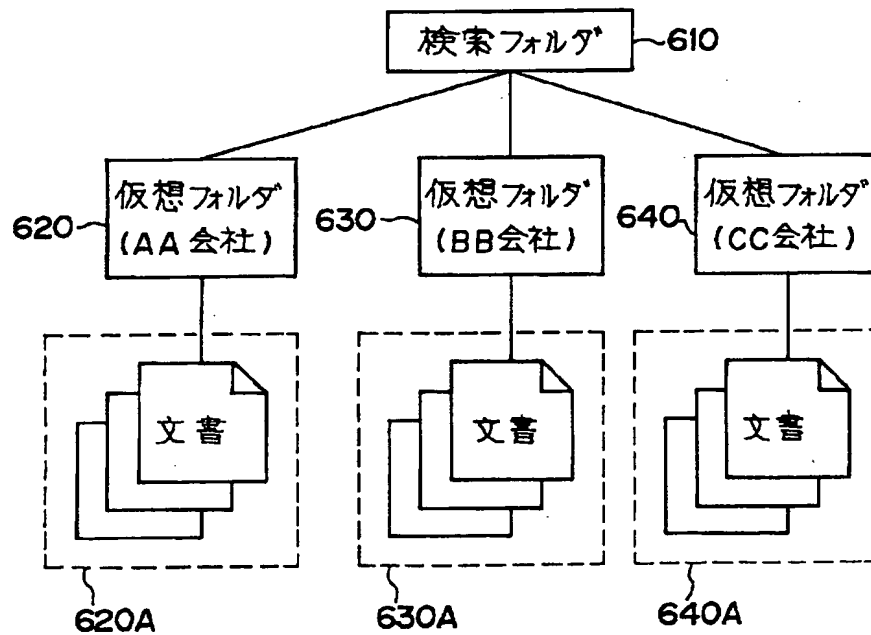
AND OR

分類属性 : 公開年

【図5】



【図6】



【図8】

800 プロパティシート

Profile : /動的分類フォルダ

検索対象 :	<input type="text" value="/日本特許フォルダ"/>	510
検索条件 :	<input type="text" value="(国際特許分類=G06F15/20)"/>	520
	<input type="checkbox"/> AND <input checked="" type="checkbox"/> OR	
分類属性 :	<input type="text" value="出願人&lt;br/&gt;公開年"/>	530

【図10】

動的な分類フォルダ

1000

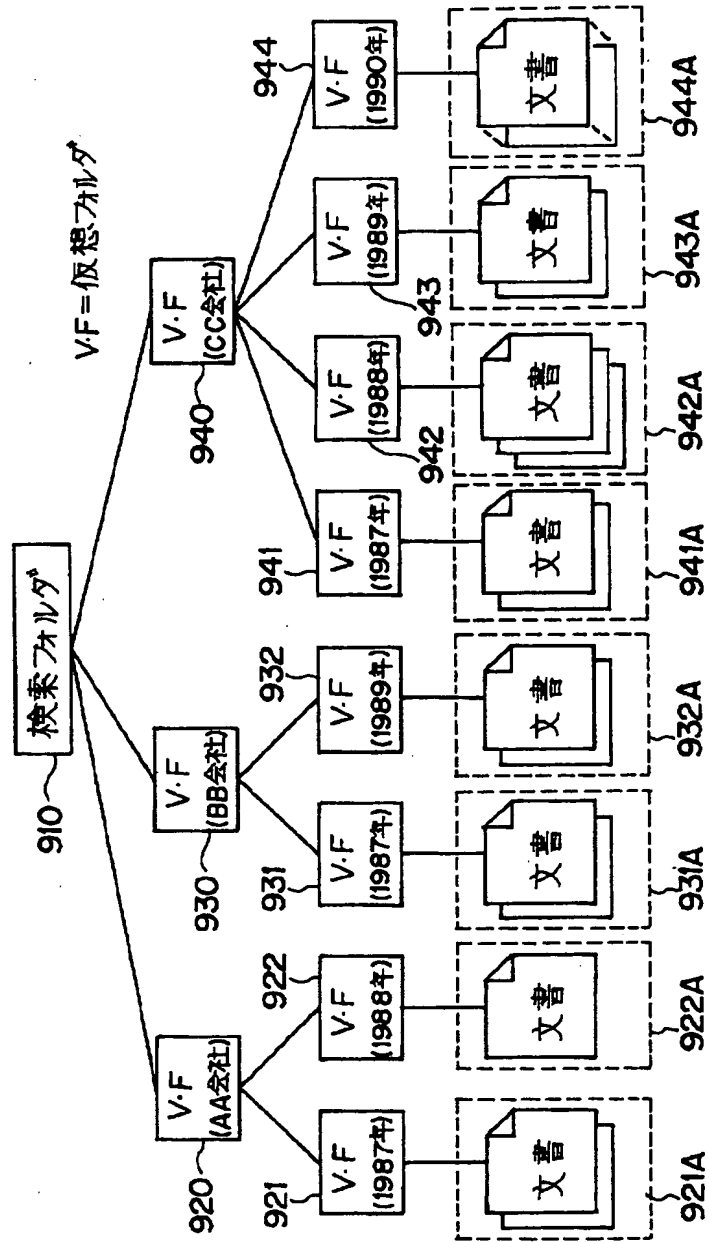
AA会社

BB会社

CC会社

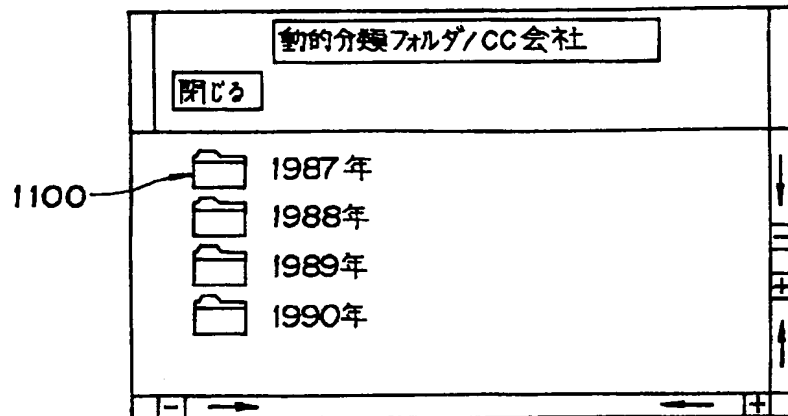
1000 points to the list area.

【図9】





【図11】



【図13】

## 1300 プロパティシート

Profile : /動的分類フォルダ/CC会社/1987年

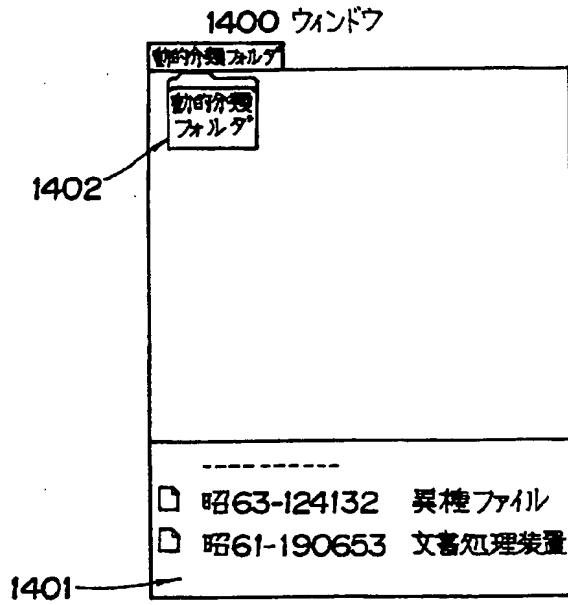
検索対象 : /動的分類フォルダ / CC会社

検索条件 : (公開年 = 1987年)

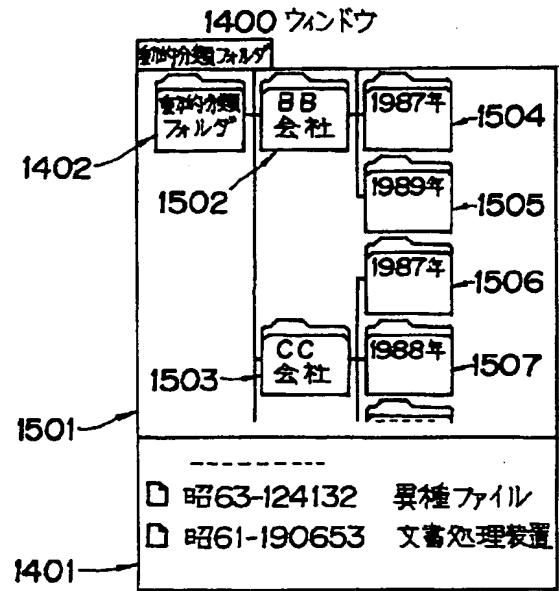
AND OR

分類属性 :

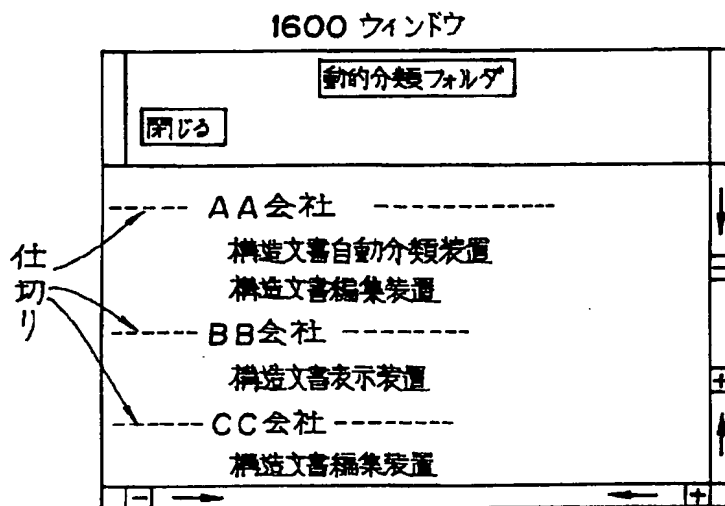
【図14】



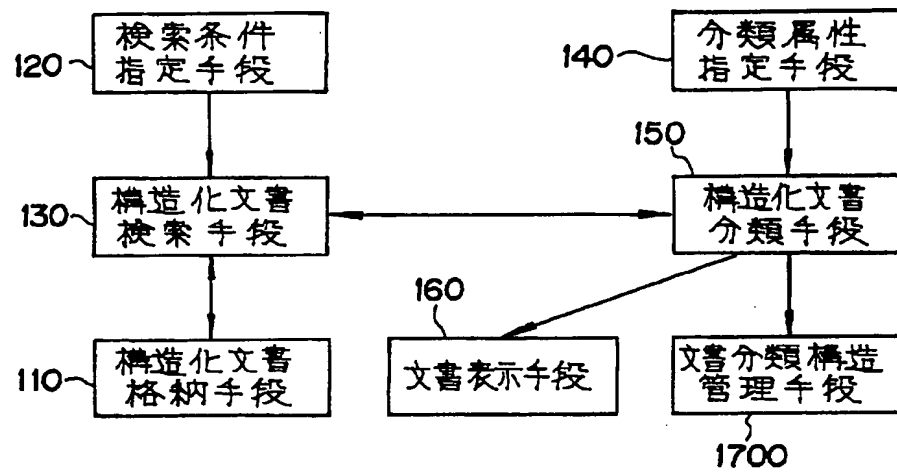
【図15】



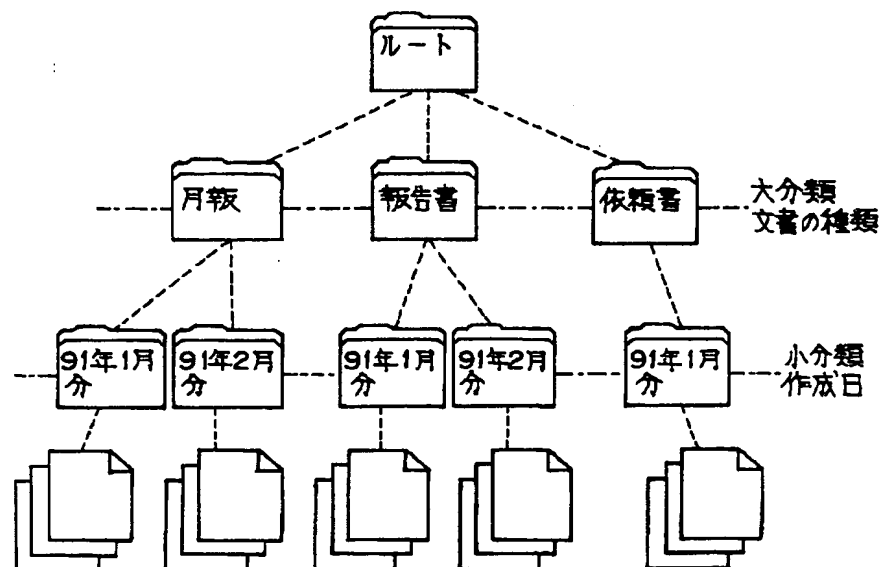
【図16】



【図17】



【図18】



【図19】

